Transferable Force Fields from Experimental Scattering Data with Machine Learning Assisted Structure Refinement

Brennon L. Shanks,[†] Jeffrey J. Potoff,[‡] and Michael P. Hoepfner[*,†]

[†]*Department of Chemical Engineering, University of Utah, Salt Lake City, UT*

[‡]*Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI*

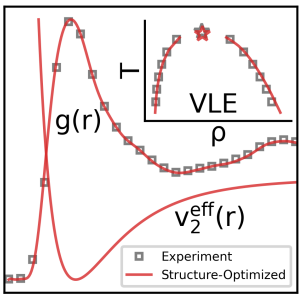
E-mail: michael.hoepfner@utah.edu

Phone: +1 (801)-581-3504

Abstract

Deriving transferable pair potentials from experimental neutron and X-ray scattering measurements has been a longstanding challenge in condensed matter physics. State-of-the-art scattering analysis techniques estimate real-space microstructure from reciprocal-space total scattering data by refining pair potentials to obtain agreement between simulated and experimental results. Prior attempts to apply these potentials with molecular simulations have revealed inaccurate predictions of thermodynamic fluid properties. In this letter, a machine learning assisted structure-inversion method applied to neutron scattering patterns of the noble gases (Ne, Ar, Kr, and Xe) is shown to recover transferable pair potentials that accurately reproduce both microstructure and vapor-liquid equilibria from the triple to critical point. Therefore, it is concluded that a single neutron scattering measurement is sufficient to predict macroscopic thermodynamic properties over a wide range of states and provide novel insight into local atomic forces in dense monoatomic systems.

TOC Graphic



Advances in neutron and X-ray scattering analysis have significantly furthered our understanding of self-assembly and dynamic transport properties in dense fluid systems.[1,2] Scattering analysis is therefore an important and necessary component in the development and validation of atomistic force fields aimed at predicting both micro- and macroscopic thermodynamic properties over a wide range of states. However, strikingly contradictory predictions between experimental and simulated microstructures have been reported in relatively simple systems, including monoatomic liquid metals,[3] aromatic hydrocarbons,[4] and water.[5] Given the proliferation of accessible neutron and X-ray scattering instrumentation, advances in computational analysis, and development of machine learning approaches, it is relevant to revisit whether scattering data can improve force fields for fluid property predictions and provide insight into local atomic forces.

One approach to benchmark force fields to scattering data is to calculate the underlying interatomic potentials from the experimental pair correlation functions, the so-called *inverse problem* of statistical mechanics. A number of well-established inversion techniques have been proposed, including Ornstein-Zernike (OZ) integral relation methods,[6–10] Yvon-Born-Green (YBG) theories,[11–15] Schommer's algorithm,[16] hypernetted chain methods,[17,18] the generalized Lyubartsev–Laaksonen approach,[19–21] empirical potential structure refinement (EPSR),[22] and a neural network.[23] However, there is little evidence that interatomic potentials obtained from these techniques can reliably predict thermodynamic behavior for real liquids. For example, Soper showed that O-O and H-H site-site interatomic potentials derived from EPSR applied to scattering data of liquid water predict a 4 times more negative excess internal energy compared to the experimental value[22] and later concluded that EPSR cannot be used to derive a reliable set of site-site pair potentials for a given system.[24] A recent scattering study on supercritical krypton found rapid short-range oscillations in EPSR-derived interatomic potentials that led the authors to conclude that augmentation of the EPSR algorithm is required to obtain a more accurate representation of the real physical system.[25] Additionally, the remaining studies on structure-inversion of real liquids reported no validation of the interatomic potentials to predict fluid properties.[26–30] Notably, in a review of structure-inversion methods it is opined that the general purpose of these techniques is not

to derive or evaluate interatomic potentials, but rather to determine molecular configurations that are consistent with the scattering data.[31] Therefore, it remains to be shown if scattering derived potentials can predict atomic trajectories consistent with experimental scattering measurements while also accurately modeling other thermodynamic properties.

The atomic length scale probed by experimental scattering measurements also confers an additional opportunity, specifically whether it is possible to learn details of the local interactions independent of assumptions on a specific model potential form (e.g., 12-6 Lennard-Jones). For example, the rate of short-range repulsive decay indicates the propensity of an atom to deform in a collision, such that relaxation from an infinitely steep potential wall to a finite exponential or power-law decay represents the transition from hard- to soft-particle collision dynamics. The approximate collision diameter may be estimated by the radial position where the potential energy intersects zero, and the pairwise radial separation of zero force describes the effective dispersion energy. Provided structure-optimized potentials demonstrate the ability to predict emergent thermodynamic properties, their application provides a bridge between local atomic physics and continuum behavior.

In this letter, force fields were determined for four noble liquids (Ne, Ar, Kr, Xe) using a machine-learning augmented Schommer's algorithm, referred to as structure-optimized potential refinement (SOPR), to refine pair potentials and obtain convergence between simulated and experimental pair distribution functions. Modifications to an initial reference potential are informed by numerical implementation of the point-wise Henderson's inverse theorem and augmented via Gaussian process regression with a squared-exponential kernel function described in Equations (11) and (14), respectively. The structure-optimized potentials predict excellent representations of both the experimental pair distribution functions (Figure 1) and saturated vapor-liquid fluid properties. Consequently, structure-optimized potentials are validated using experimentally-consistent observations on both the micro- and macroscopic length scales, motivating the analysis of specific properties of the generated potentials. Additionally, the monoatomic structure and spherical symmetry of the noble gas system facilitates the comparison of the structure-optimized potentials

(Figure 2) to reference *ab initio* potentials obtained in the low-density state from coupled cluster theory,[32–35] referred to as reference quantum dimer potentials. This comparison reveals state-dependent changes of many-body forces present in the experimental systems that were collected at states with varying reduced temperatures (T_r) relative to the critical point.

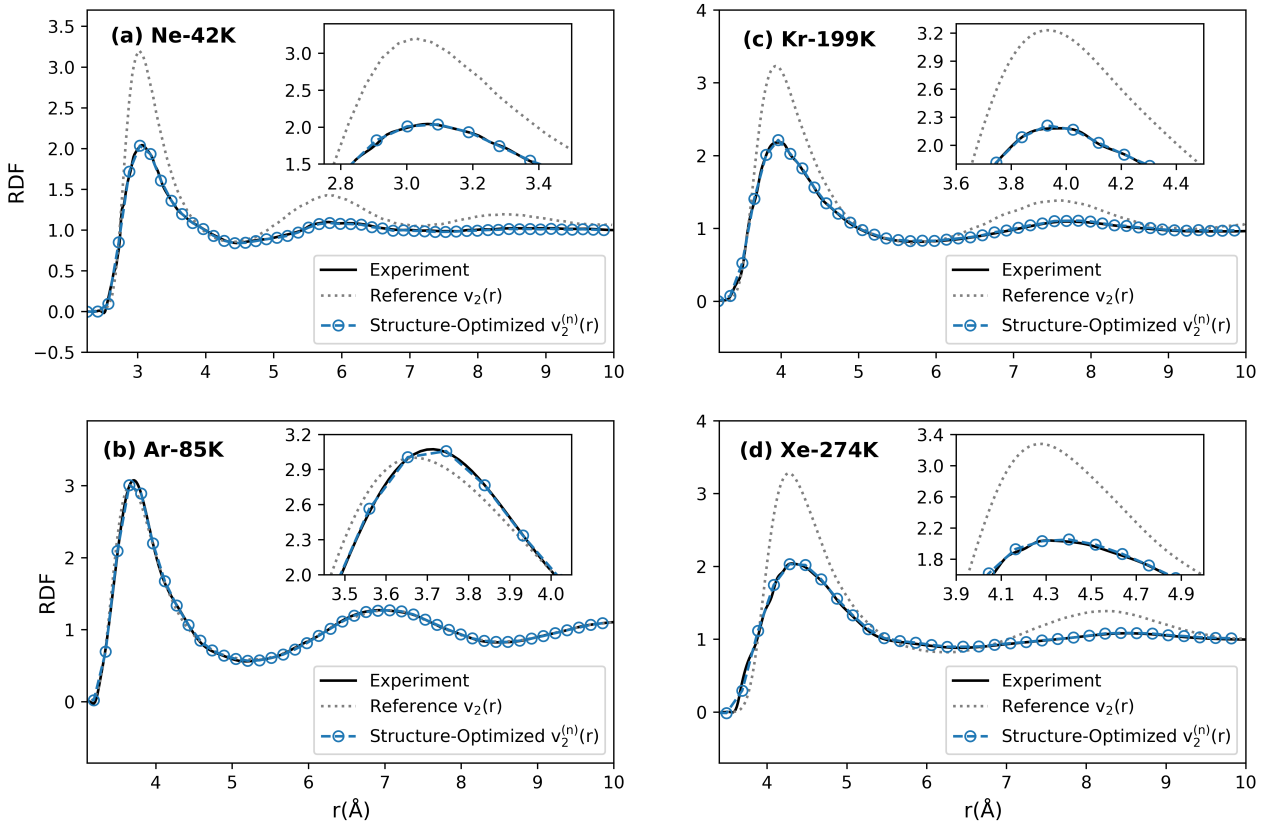


Figure 1: Reference potential radial distribution functions (grey dotted line) compared to the converged simulated radial distribution functions (blue circles) and experimental radial distribution functions (black line). Inset figures show the first solvation shell of the radial distribution function.

The structure-optimized potentials collected for fluids near their critical point (Ne-42K, Kr-199K, Xe-274K at $T_r = 0.95$) exhibit softer repulsive decay, insignificant change to the collision diameter, and a substantial reduction in dispersion energy with respect to reference quantum dimer potentials. Thus, the ensemble averaged many-body behavior near the critical point results in softer particle collisions with decreased particle attraction. Near the triple point (Ar-85K at $T_r = 0.56$), structure-optimized potentials show no significant change in the repulsive exponent, a 1.5 % increase in collision diameter, and a reduction in dispersion energy (Figure 2 (b)) compared to the

quantum dimer potential. Many-body effects therefore had a negligible effect on the particle stiffness while decreasing particle attraction near the triple point. The observation that the dispersion energy correction was relatively smaller for the near triple point potential compared to the near critical point potentials (Table 3) suggests that the dispersion energy is a function of the thermodynamic state, which is discussed in the context of temperature-dependent many-body effects later.

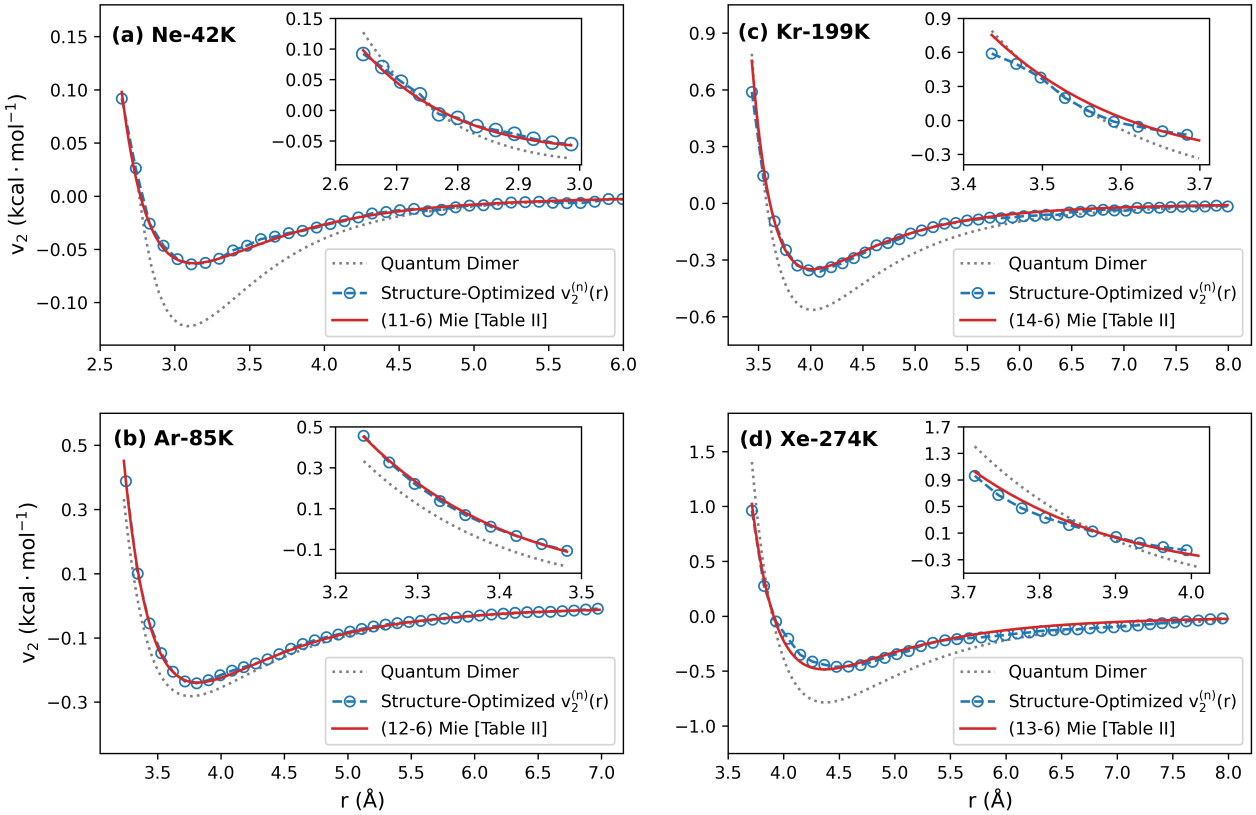


Figure 2: Tabulated structure-optimized potentials (blue) and $(\lambda - 6)$ Mie potentials determined with Bayesian regression (red) are shown with reference quantum dimer potentials (grey). Inset figures show the short-range repulsive region of the corresponding structure-optimized potential.

It is instructive to compare the structure-optimized potentials to widely employed transferable pair potential functions, such as the $(\lambda - 6)$ Mie potential,

$$v_2^{Mie}(r) = \frac{\lambda}{\lambda - 6}\left(\frac{\lambda}{6}\right)^{\frac{6}{\lambda-6}} \varepsilon \left[\left(\frac{\sigma}{r}\right)^{\lambda} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1}$$

where λ is the short-range repulsion exponent, σ is the collision diameter (Å), and ε is the dis-

persion energy (kcal $\cdot$ mol^{-1}).[36] The $(\lambda - 6)$ Mie potential offers increased flexibility over the standard $(12-6)$ Lennard-Jones potential since the repulsion exponent may be varied to produce a wider array of potential shapes. Structure-optimized potentials were fit to the $(\lambda - 6)$ Mie function via Bayesian regression and plotted as red lines in Figure 2. Note that the excellent quality-of-fit of structure-optimized potentials to the $(\lambda - 6)$ Mie function indicates that the fitted parameters (listed in Table 1) can closely approximate the thermodynamic predictions of the tabulated structure-optimized potentials.

Table 1: Summary of $(\lambda - 6)$ Mie potential parameters determined by Bayesian linear regression and modifications to the reference quantum dimer potentials in terms of the (σ, ε) parameterization. $\Delta\sigma$ and $\Delta\varepsilon$ are shown as percent deviations from the *ab initio* parameter values given in Table 3.

Element	λ	σ (Å)	ε (kcal/mol)	$\Delta\sigma$ (%)	$\Delta\varepsilon$ (%)
Ne	11	2.77	0.063	0.31	-48.4
Ar	12	3.40	0.239	1.50	-16.7
Kr	14	3.58	0.359	-0.08	-38.3
Xe	13	3.91	0.484	0.51	-40.3

Transferability of the potentials was assessed by performing vapor-liquid equilibrium (VLE) calculations from the triple to critical point using histogram-reweighting grand canonical Monte Carlo (GCMC) simulations in the GPU-Optimized Monte Carlo (GOMC) simulation package[37] (see Supporting Information). Figure 3 shows vapor and liquid densities for structure-optimized potentials fit to $(\lambda - 6)$ Mie potentials (red triangles) compared with experimental data (black lines) compiled from the National Institute of Standards and Technology (NIST).[38] The Ne-42K structure-optimized force field predicts liquid densities within 0.1-2.5% relative error between 30-40K, on par with the top-performing Lennard-Jones force field from Vrabec *et al.*[39] and outperforming the next closest model[40] by as much as 10%. The Ar-85K, Kr-199K, and Xe-274K force fields are less accurate, with liquid density relative errors of 0.2-5% (85-140K), 6.2-10.1% (120-180K) and 4.7-8.4% (190-260K), respectively. Simulated critical points determined with the Ising-type critical point scaling law[41] and law of rectilinear diameters[42] are provided in Table 2.

Table 2: Simulated critical temperatures (T_C^{sim}) and densities (ρ_C^{sim}) with statistical uncertainty calculated from 5 independent GCMC simulations. Percent error between simulated and experimental critical temperature and density are also shown.

Element	T_C^{sim} (K)	T_C^{err} (%)	ρ_C^{sim} (kg/m^3)	ρ_C^{err} (%)
Ne	43.84 $\pm$ 0.06	-1.33	488.4 $\pm$ 1.08	0.91
Ar	154.27 $\pm$ 0.16	2.40	528.9 $\pm$ 0.84	-1.32
Kr	216.58 $\pm$ 0.60	3.36	952.7 $\pm$ 3.62	5.04
Xe	300.99 $\pm$ 0.28	3.98	1142.9 $\pm$ 1.76	4.85

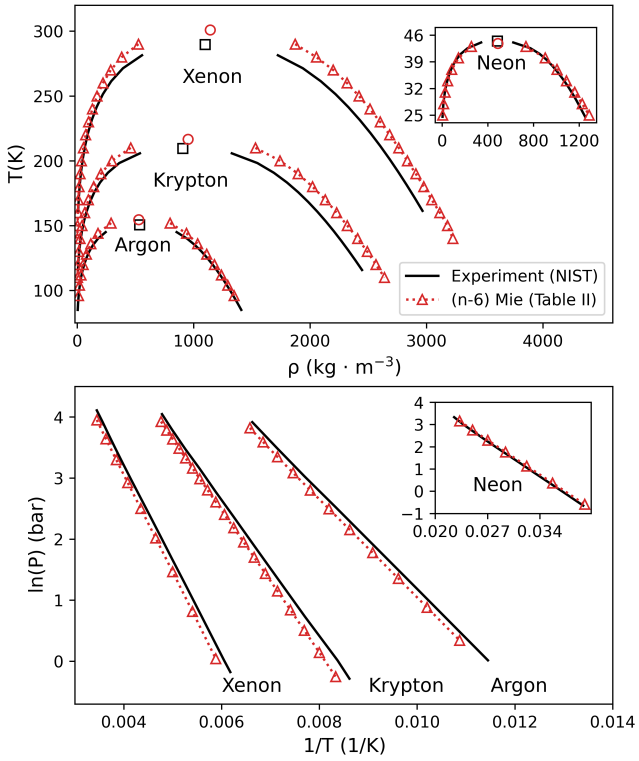


Figure 3: (Top) Simulated phase coexistence curves (red triangles) are shown with experimental phase data (black lines). Simulated and experimental critical points are given by the open red circle and open black square, respectively. (Bottom) Clausius-Clapeyron plots of the simulated pressure (red triangles) compared to experimental pressures (black lines).

A recently developed series of $(\lambda - 6)$ Mie force fields benchmarked to noble gas vapor-liquid equilibrium (VLE) provides an excellent comparison to the structure-optimized force fields proposed in this work. In general, both force fields predict similar repulsion exponents (λ) and dispersion energies (ε). One interesting observation is that both the structure and VLE-optimized force fields predict an increase in the repulsive exponent with increasing atomic weight. Mick *et al.*[43]

demonstrated that varying the repulsive exponent improved the simultaneous prediction of saturated densities and vapor pressures, supporting the conclusion that static structure is sensitive to subtle variations in the pair potential. The most consequential difference between the two models is that the structure-optimized force field gives systematically lower predictions for the collision diameter (σ), causing the simulated liquid density to be overestimated compared to the experimental value. It is notable that the reported differences in the collision diameter (0.01-0.05 Å) are approximately one order of magnitude smaller than the real-space resolution of the experimental diffraction data (0.4-0.7 Å). Modern, high-flux scattering instruments can achieve real-space resolution of approximately 0.05 Å, on the same order as the error in the collision diameter, suggesting that more accurate potentials may be determined from repeating neutron scattering measurements on noble gases with modern instruments.[44]

The observation that structure-optimized potentials can estimate vapor-liquid coexistence behavior from a single neutron scattering measurement suggests that structure-inversion may be a promising approach to develop force fields for materials where experimental phase behavior is absent or impractical to obtain. The phase behavior and criticality of liquid uranium (U) is one of many important and unresolved examples relating to nuclear reactor design and safety analysis. Neutron diffraction data on solid β-U exists to temperatures as high as 1045K[45] and X-ray diffraction patterns of γ-U close to the melting point (1405K) are well-characterized,[46] but the phase coexistence line and critical point remain unknown, with critical temperature predictions ranging from 5000-13000K.[47] However, our results suggest that scattering measurements of liquid U may enable estimation of vapor-liquid phase coexistence via GCMC simulations with a structure-optimized embedded atom model or set of state dependent structure-optimized potentials.

Structure-inversion may also enable quantification of liquid state many-body effects. Quantum calculations on Kr-Ar-Ar and Ar-Ar-Kr trimers have revealed that noble gases experience two important many-body effects: (1) 3-body exchange repulsion and (2) the exchange/dispersion quadropole induced dipole.[48] The averaged pairwise influence of these many-body interactions decreases the short-range repulsion exponent and the dispersion energy,[49,50] which is in agree-

ment with the behavior observed in the structure-optimized potentials obtained near the critical point. Self-consistent field calculations of electron distributions in Ar clusters have shown that the electron cloud is compressed at higher densities, and that this compression reduces the probability of exchange repulsion.[51] Additionally, experimental results from collision-induced depolarized light scattering on compressed H_2 demonstrated that exchange-dependent many-body interactions become more prominent with increasing temperature.[52] It is therefore expected that many-body effects in noble gases should be less dominant near the triple point, explaining why the short-range decay rate for the Ar-85K structure-optimized potential was unchanged and the well-depth correction smaller than in the near critical point states. This conclusion is also supported by analyzing trends in the structural many-body correction for each fluid at near triple and critical point conditions (see Supporting Information). Further analysis with quantum mechanical and explicit 3-body dispersion models, such as the Axilrod-Teller potential,[53–55] are reserved for future study on high-resolution scattering data sets obtained with state-of-the-art neutron techniques.

We demonstrate that transferable pair potentials can be reconstructed from a single neutron scattering measurement for monoatomic liquids, and further; that structure-inversion techniques have fundamental and interdisciplinary applications bridging experimental scattering, molecular simulation, and quantum mechanics. Of particular interest is the prediction of thermodynamic properties at extreme conditions, such as high temperature and pressure materials, molten salts, and liquid metals. The inclusion of experimental diffraction results for optimizing effective pair potentials may also facilitate improvements to local structure predictions for fluid mixtures and molecular liquids. However, incoherent and inelastic scattering corrections,[56] as well as non-uniqueness of the partial structure factor decomposition, will need to be addressed to extend the presented techniques to complex liquids. Finally, the methods presented in this letter may be applied to benchmark force fields for coarse-grained simulations, which has seen a growing interest in structure-inversion techniques.[57,58]

Theory and Computational Methods

The following section provides the relevant definitions, statistical mechanics, and necessary computational details of the proposed machine learning assisted structure refinement method. First, the microstructure is considered as the local atomic density correlation and is formalized by counting the number of atomic neighbors as a function of position with respect to a reference atom and taking the ensemble average,

$$g(\mathbf{r}) = \frac{1}{\rho}\left\langle \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta^3(\mathbf{r}-\mathbf{r}_j+\mathbf{r}_i)\right\rangle \tag{2}$$

where $g(\mathbf{r})$ is referred to as the radial distribution function, δ^3 is the three-dimensional Dirac delta function, ρ is the thermodynamic density and N is the total number of particles in the system. The radial distribution function and pair correlation function, $h(\mathbf{r})$, are related by, $h(\mathbf{r}) = g(\mathbf{r}) - 1$. Due to the lack of long-range order in liquids, the isotropically-averaged radial distribution function is related to the experimentally observed structure factor, $S(Q)$, which for a monoatomic liquid is given by,

$$S(Q) = 1 + \frac{4\pi}{Q\rho}\frac{\langle b\rangle^2}{\langle |f(Q)|^2\rangle}\int_0^\infty r[g(r)-1]\sin(Qr)dr \tag{3}$$

where Q is the momentum transfer, b is the scattering length density, $f(Q)$ is the form factor, and ρ is the atomic number density.[59]

The potential energy can be written as a sum of n-body potential energy terms such that,

$$U(\mathbf{r}) = \overbrace{\sum_{i}^{N} v_1(\mathbf{r}_i)}^{\text{external field}} + \overbrace{\sum_{i=1}^{N}\sum_{j\neq i}^{N} v_2(\mathbf{r}_{ij})}^{\text{two-body}} + \overbrace{\sum_{i}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j}^{N} v_3(\mathbf{r}_{ijk}) + \ldots}^{\text{many-body}} \tag{4}$$

where $v_p(\mathbf{r}_{1,\ldots,p})$ is a position dependent function that assigns a potential energy to a subset containing $p \leq N$ atoms for a given configuration $\mathbf{r}_{i,\ldots,p}$.[60] We further simplify this expression by neglecting the external field contribution ($p = 1$) and averaging higher-order many-body terms

($p \geq 3$) into a state-dependent pair term,

$$U(\mathbf{r};\rho,T) = \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[v_2(\mathbf{r}_{ij}) + v_2^m(\mathbf{r}_{ij};\rho,T)\right] \tag{5}$$

such that $v_2^m(\mathbf{r}_{ij};\rho,T)$ is explicitly dependent on the atomic positions and implicitly dependent on the physical state (temperature, density, etc). The bracketed quantity in Equation (5) is defined as the effective pair potential,

$$v_2^{eff}(\mathbf{r}_{ij};\rho,T) = v_2(\mathbf{r}_{ij}) + v_2^m(\mathbf{r}_{ij};\rho,T) \tag{6}$$

which cannot be determined exactly for a state-dependent ensemble[61] but can be optimized to reproduce a set of experimentally observed thermodynamic properties, such as structure,[62] heat of vaporization,[63] or vapor-liquid equilibrium.[37] The pair potential defined in Equation (6) is the most common non-bonded term in the Hamiltonian of classical force fields and is typically modeled as a hard-particle, Lennard-Jones, $(\lambda - 6)$ Mie, Buckingham, or Yukawa potential.

Pairwise additivity imposes important theoretical constraints on the relationship between the potential energy and pair correlation function. Henderson proved that for pairwise additive, constant density ensembles that there exists a one-to-one map between the effective pair potential and the radial distribution function up to an additive constant.[64–66] In monoatomic liquids with spherical symmetry, the structure-potential uniqueness theorem on a finite radial interval $[r', r'']$ such that $r'' > r'$ and $r', r'' \in \mathbb{R}_0^+$, is equivalent to,

$$\int_{r'}^{r''} \Delta v_2(r)\Delta g(r)\, dr \leq 0 \tag{7}$$

where $\Delta v(r)$ and $\Delta g(r)$ are the difference between a model (M) and target (T) pair potential and radial distribution function, respectively.[67]

$$\begin{aligned} \Delta v_2(r) &= v_2^M(r) - v_2^T(r) \\ \Delta g(r) &= g^M(r) - g^T(r) \end{aligned} \tag{8}$$

Note that the structure-potential uniqueness theorem in the form of Equation (7) is written in terms of the r-coordinate only due to spherical symmetry. Initially, Equation (7) appears uninformative since the inequality prevents direct calculation of the target potential at any r value. The situation is amenable under the assumption that the integrand is continuous and differentiable, so that Equation (7) can be rewritten using the mean value theorem,

$$(r'' - r') \left\langle \Delta v_2(r) \Delta g(r) \right\rangle_{r'}^{r''} \leq 0 \tag{9}$$

where the bracketed quantity represents the average of $\Delta v_2(r)\Delta g(r)$ over finite interval $[r', r'']$. Notice that for this inequality to be satisfied in the limit $(r'' - r') \to 0$, it must hold at any point $r_o \in r$ so that,

$$\Delta v_2(r_o) \Delta g(r_o) \leq 0 \tag{10}$$

which is true only when $\Delta v_2(r_o) \neq 0$ and $\Delta g(r_o) \neq 0$. The practicality of this point-wise structure-potential uniqueness theorem is now clear, since Equation (10) prescribes what direction that an initial guess for the model potential should be corrected given the difference between the model and target experimental radial distribution function at any point $r_o \in r$; namely, by decreasing the potential if $\Delta g(r_o)$ is negative and increasing the potential if $\Delta g(r_o)$ is positive. While Henderson's structure-potential uniqueness condition has been implemented previously to obtain empirical estimates of pair potential functions in Schommer's algorithm and EPSR, this derivation demonstrates the validity of its use at an arbitrary point without the potential of mean force approximation $g(r) = \exp[-\beta v_2^{eff}(r_{ij}; \rho, T)]$ where $\beta = \frac{1}{k_B T}$, which only holds in the dilute limit.[68]

The structure-potential uniqueness condition is implemented via iterative refinement of a reference potential, $v_2^0(r_i)$, with an energy scaled, continuous sum of the radial distribution function

error such that,

$$v_2^{(n)'}(r_i) = v_2^0(r_i) + \gamma\beta^{-1}\sum_n \Delta g^{(n)'}(r_i) \tag{11}$$

where i is the radial index of the tabulated potential, $v_2^{(n)'}(r_i)$ is the predictor estimated pair potential at iteration n, β is the inverse thermal energy $(k_B T)^{-1}$, and $0 < \gamma \leq 1$ is an empirical scaling constant to dampen the potential correction. Comparing the refinement Equation (11) to Equation (6), it is clear that if $v_2^0(r_i)$ is selected as the quantum dimer potential that $v_2^{(n)'}(r_i)$ is the estimated effective pair potential and $\gamma\beta^{-1}\sum_n \Delta g^{(n)'}(r_i)$ is the pair averaged many-body term. Note that the prime notation in $v_2^{(n)'}(r_i)$ denotes that the pair potential is the predictor estimate before smoothing and treatment of numerical and experimental uncertainty.

In a standard Schommer's algorithm, the potential predicted by Equation (11) is passed to the next iteration without smoothing or uncertainty quantification, which has been shown to reduce the methods robustness.[17,31] Here a squared-exponential kernel Gaussian process (GP) is applied to the predictor estimate to account for numerical fluctuations arising from the molecular dynamics simulations as well as systematic over-fitting to uncertain experimental data. A GP is a non-parametric stochastic process, equivalent to an infinitely wide neural network of a single layer, that generalizes the concept of probability distributions to functions.[69] In this implementation, the GP takes the potential estimated by Equation (11) as an input and returns a Gaussian probability distribution of continuous and infinitely differentiable functions fitting the predictor estimate.[70] Thus, a GP acts as an uncertainty propagator and smoothing function that, by nature of its Gaussian form, inherits an analytical Fourier transform that equivalently represents the data in real- or reciprocal-space without introducing significant truncation error.[71] Parallel techniques to enhance the accuracy of Fourier transforms in inverse problems, such as fitting structure factors to Poisson series expansions implemented in the EPSR and *Dissolve* packages and Tóth's Gauss-Newton parameterization and Golay–Savitzky smoothing,[20,21] can therefore be replaced with GP regression. Notably, SE-GP regression can be integrated into any existing iterative predictor-corrector without modifications to the base algorithm.

The predicted structure-optimized potential is then expressed as a k-multivariate normal distribution ($\mathscr{N}$) of random variables such that,

$$v_2^{(n)'}(r_1), v_2^{(n)'}(r_2), ...v_2^{(n)'}(r_k) \sim \mathscr{N}(\mu(\mathbf{r}), \mathbf{K}(r_i, r_j)) \tag{12}$$

where $r_1 < r_2 < ... < r_k$ are the radii positions for the potential, k is the number of points in the structure-optimized potential, $\mu(\mathbf{r})$ is a mean function, and $\mathbf{K}(r_i, r_j)$ is a squared-exponential covariance function (or kernel) describing the relatedness of observations $v_2^{(n)'}(r_i)$ on $v_2^{(n)'}(r_j)$. Here the squared-exponential kernel is applied,

$$\mathbf{K}(r_i, r_j) = \bar{\sigma}^2 e^{-\frac{(r_i - r_j)^2}{2\ell^2}} + \delta_{ij}\sigma_{noise}^2 \tag{13}$$

where $\bar{\sigma}^2$ is the expected variance of the interatomic potential, ℓ is the correlation length and σ_{noise}^2 is the variance due to numerical latent effects. Notice that if the distance between two points r_i and r_j is small, $\exp(r_i - r_j)^2/2\ell^2$ approaches unity and $v_2^{(n)'}(r_i)$ and $v_2^{(n)'}(r_j)$ are strongly correlated. As the distance between r_i and r_j increases, $\exp(r_i - r_j)^2/2\ell^2$ vanishes such that $v_2^{(n)'}(r_i)$ and $v_2^{(n)'}(r_j)$ are uncorrelated. The hyperparameters ($\bar{\sigma}$, ℓ, σ_{noise}) are optimized by maximizing the marginal likelihood (model evidence) $p(v_2^{(n)'}(r)|r, \bar{\sigma}, \ell, \sigma_{noise})$.

Regression of $v_2^{(n)'}(r)$ over an arbitrary set of radii $r' = \{r_i'\} \ni r_1 \leq r_1' < r_2' < ... < r_m' \leq r_k$ is equal to the mean of the k-variate normal distribution,

$$v_2^{(n)}(r') = [\mathbf{K}_{r',r}^T - \sigma_{noise}^2\mathbf{I}]\mathbf{K}_{r,r}^{-1} v_2^{(n)'}(r) \tag{14}$$

where $v_2^{(n)}(r')$ is the final structure-optimized potential at iteration n and $\mathbf{K}_{r',r}$ is the squared-exponential covariance matrix between coordinate representations r' and r. Figure 4 shows that GP regression smooths numerical artifacts in the interatomic force when the length scale hyperparameter is on the order of $\ell \sim 1$ Å. A detailed comparison between a standard and GP assisted Schommer's algorithm is provided in the Supporting Information.

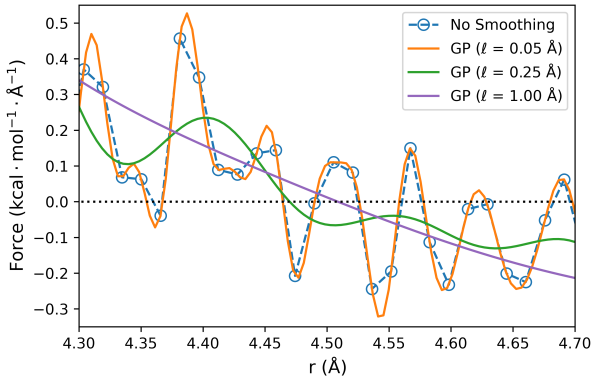


Figure 4: The force calculated from the Xe-274K tabulated potential (blue circles) compared to a non-parametric GP regressed potential at three different length-scale parameters (ℓ = 0.05, 0.25, 1.00 Å) (orange, green, and purple line).

The GP regressed structure-optimized potential is then applied in the molecular simulation corrector to calculate a simulated radial distribution function, $g^{(n)}(r')$. The molecular simulation corrector is a Canonical (NVT) molecular dynamics simulation performed in HOOMD-Blue.[72] MD simulations were initiated with a 500 atom fcc-lattice at the experimental density and equilibrated with Langevin dynamics for 5×10^3 timesteps ($dt = 5$ femtoseconds). Tabulated structure-optimized potentials were truncated at $3r_{vdW}$ with analytical tail corrections, and simulated radial distribution functions were calculated with MDAnalysis[73,74] from 1×10^5 timestep trajectories sampled at 100 timestep intervals. Convergence is checked against the average squared error between the simulated and experimental radial distribution function such that $\langle[\Delta g^{(n)}(r')]^2\rangle < 10^{-4}$, which generally is satisfied within 5-10 iterations at scaling constant $\gamma = 0.2$.

Table 3: Reduced temperature ($T_r = T/T_c$) and atomic density (ρ) are listed for the neutron scattering experimental conditions. Reference van der Waal radii (r_{vdW}) are used to define pair potential truncation in molecular dynamics simulations.[75] σ_{ai} is defined as the radius where the quantum dimer pair potential transitions from positive to negative potential energy and ε_{ai} is the potential minimum.

Element	T_r	ρ (1/Å^3)	r_{vdW} (Å)	σ_{ai} (Å)	ε_{ai} (kcal/mol)
Ne	0.95	0.02477	2.91	2.76	0.122
Ar	0.56	0.02125	3.55	3.35	0.287
Kr	0.95	0.01187	3.82	3.58	0.582
Xe	0.95	0.00881	4.08	3.89	0.811

Structure-inversion was initiated with a target experimental radial distribution function and a reference (or model) pair potential, $v_2^0(r_i)$. Experimental radial distribution data determined with elastic neutron scattering[76–78] were compiled at the thermodynamic state conditions listed in Table 3. Reference quantum dimer potentials were obtained via couple-cluster theory/t-aug-cc-pV6Z quality basis sets with spin-orbit relativistic corrections. In practice, any of the numerous existing pair potentials for the noble gases may be applied as a reference potential with equivalent outcomes for the structure-optimized potential (see Supporting Information). However, selecting the quantum dimer pair potential as the reference guarantees that the structure-optimized refinement correction is equal to the pairwise many-body contribution to the effective pair potential, $v_2^m(\mathbf{r}_{ij};\rho,T) = \gamma\beta^{-1}\sum_n \Delta g^{(n)}(r_i')$.

Acknowledgement

This study is supported by the National Science Foundation Award No. CBET-1847340. The authors thank Sean Thomas Smith, Philip J. Smith and Harry W. Sullivan for fruitful discussions on the implementation of Gaussian process regression and Terry Arthur Ring for generously reviewing the manuscript prior to submission.

Supporting Information Available

The following files are available free of charge as supporting information for the results presented in this manuscript.

- si.pdf: Grand canonical Monte Carlo simulation details, data tables of vapor-liquid phase equilibrium data, liquid density error compared with other force fields, convergence stability analysis, a demonstration of a standard versus machine learning augmented Schommer's algorithm, and potentials derived for Ne, Kr, and Xe at both the near triple and critical point conditions are provided.

- data.zip: Compressed .zip file containing .txt files of experimental, reference, and final simulated radial distribution functions and interatomic potentials.

Example code can be accessed at: https://github.com/hoepfnergroup/SOPR.

References

(1) Glotzer, S. C.; Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nat. Mater.* **2007**, *6*, 557–562.

(2) Wu, L.; Willis, J. J.; McKay, I. S.; Diroll, B. T.; Qin, J.; Cargnello, M.; Tassone, C. J. High-temperature crystallization of nanocrystals into three-dimensional superlattices. *Nature* **2017**, *548*, 197–201.

(3) Itami, T.; Munejiri, S.; Masaki, T.; Aoki, H.; Ishii, Y.; Kamiyama, T.; Senda, Y.; Shimojo, F.; Hoshino, K. Structure of liquid Sn over a wide temperature range from neutron scattering experiments and first-principles molecular dynamics simulation: A comparison to liquid Pb. *Phys. Rev. B* **2003**, *67*, 064201.

(4) F. Headen, T.; L. Cullen, P.; Patel, R.; Taylor, A.; T. Skipper, N. The structures of liquid pyridine and naphthalene: the effects of heteroatoms and core size on aromatic interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 2704–2715.

(5) Amann-Winkel, K.; Bellissent-Funel, M.-C.; Bove, L. E.; Loerting, T.; Nilsson, A.; Paciaroni, A.; Schlesinger, D.; Skinner, L. X-ray and Neutron Scattering of Water. *Chem. Rev.* **2016**, *116*, 7570–7589.

(6) Ornstein, L.; Zernike, F. Integral equation in liquid state theory. *Proc. Acad. Sci. Amsterdam.* **1914**, *17*.

(7) Percus, J. K.; Yevick, G. J. Analysis of Classical Statistical Mechanics by Means of Collective Coordinates. *Phys. Rev.* **1958**, *110*, 1–13.

(8) Percus, J. K. Approximation Methods in Classical Statistical Mechanics. *Phys. Rev. Lett.* **1962**, *8*, 462–463.

(9) Wang, Q.; Keffer, D. J.; Nicholson, D. M.; Thomas, J. B. Use of the Ornstein-Zernike Percus-Yevick equation to extract interaction potentials from pair correlation functions. *Phys. Rev. E* **2010**, *81*, 061204.

(10) A. Goodall, R. E.; A. Lee, A. Data-driven approximations to the bridge function yield improved closures for the Ornstein–Zernike equation. *Soft Matter* **2021**, *17*, 5393–5400.

(11) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*, 198104.

(12) Mullinax, J. W.; Noid, W. G. A Generalized Yvon Born Green Theory for Determining Coarse-Grained Interaction Potentials. *J. Phys. Chem. C* **2010**, *114*, 5661–5674.

(13) Ellis, C. R.; Rudzinski, J. F.; Noid, W. G. Generalized-Yvon–Born–Green Model of Toluene. *Macromol. Theory. Simul.* **2011**, *20*, 478–495.

(14) Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures. *J. Chem. Phys.* **2016**, *144*, 204124.

(15) DeLyser, M.; Noid, W. G. Bottom-up coarse-grained models for external fields and interfaces. *J. Chem. Phys.* **2020**, *153*, 224103.

(16) Schommers, W. Pair potentials in disordered many-particle systems: A study for liquid gallium. *Phys. Rev. A* **1983**, *28*, 3599–3605.

(17) Levesque, D.; Weis, J. J.; Reatto, L. Pair Interaction from Structural Data for Dense Classical Liquids. *Phys. Rev. Lett.* **1985**, *54*, 451–454.

(18) Delbary, F.; Hanke, M.; Ivanizki, D. A generalized Newton iteration for computing the solution of the inverse Henderson problem. *Inverse Probl. Sci. Eng.* **2020**, *28*, 1166–1190.

(19) Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730–3737.

(20) Tóth, G. Determination of pair-potential parameters from experimental structure factors. *J. Chem. Phys.* **2001**, *115*, 4770–4775.

(21) Tóth, G. An iterative scheme to derive pair potentials from structure factors and its application to liquid mercury. *J. Chem. Phys.* **2003**, *118*, 3949–3955.

(22) Soper, A. K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **1996**, *202*, 295–306.

(23) Tóth, G.; Király, N.; Vrabecz, A. Pair potentials from diffraction data on liquids: A neural network solution. *J. Chem. Phys.* **2005**, *123*, 174109.

(24) Soper, A. K. Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. *Mol. Phys.* **2001**, *99*, 1503–1516.

(25) Pruteanu, C. G.; Loveday, J. S.; Ackland, G. J.; Proctor, J. E. Krypton and the Fundamental Flaw of the Lennard-Jones Potential. *J. Phys. Chem. Lett.* **2022**, *13*, 8284–8289.

(26) Reatto, L.; Levesque, D.; Weis, J. J. Iterative predictor-corrector method for extraction of the pair interaction from structural data for dense classical liquids. *Phys. Rev. A* **1986**, *33*, 3451–3465.

(27) Kahl, G.; Kristufek, M. Inversion problem for dense classical liquids: The binary case. *Phys. Rev. E* **1994**, *49*, R3568–R3571.

(28) Soper, A. K. Partial structure factors from disordered materials diffraction data: An approach using empirical potential structure refinement. *Phys. Rev. B* **2005**, *72*, 104204.

(29) Youngs, T. Dissolve: next generation software for the interrogation of total scattering data by empirical potential generation. *Mol. Phys.* **2019**, *117*, 3464–3477.

(30) Zhao, Y. Structural analysis and potential extraction from diffraction data of disordered systems by least-biased feature matching. *J. Chem. Phys.* **2021**, *155*, 234501.

(31) Tóth, G. Interactions from diffraction data: historical and comprehensive overview of simulation assisted methods. *J. Phys. Condens. Matter* **2007**, *19*, 335220.

(32) Hellmann, R.; Bich, E.; Vogel, E. Ab initio potential energy curve for the neon atom pair and thermophysical properties of the dilute neon gas. I. Neon–neon interatomic potential and rovibrational spectra. *Mol. Phys.* **2008**, *106*, 133–140.

(33) Patkowski, K.; Szalewicz, K. Argon pair potential at basis set and excitation limits. *J. Chem. Phys.* **2010**, *133*, 094304.

(34) Jäger, B.; Hellmann, R.; Bich, E.; Vogel, E. State-of-the-art ab initio potential energy curve for the krypton atom pair and thermophysical properties of dilute krypton gas. *J. Chem. Phys.* **2016**, *144*, 114304.

(35) Hellmann, R.; Jäger, B.; Bich, E. State-of-the-art ab initio potential energy curve for the xenon atom pair and related spectroscopic and thermophysical properties. *J. Chem. Phys.* **2017**, *147*, 034304.

(36) Mie, G. Zur kinetischen Theorie der einatomigen Körper. *Ann. Phys.* **1903**, *316*, 657–697.

(37) Nejahi, Y.; Soroush Barhaghi, M.; Mick, J.; Jackman, B.; Rushaidat, K.; Li, Y.; Schwiebert, L.; Potoff, J. GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids. *SoftwareX* **2019**, *9*, 20–27.

(38) Lemmon, E. W. Thermophysical Properties of Fluid Systems. *NIST Chemistry Webbook : NIST Standard Reference Database Number 69* **1998**, *20899*, http://webbook.nist.gov, accessed 2021-12-15.

(39) Vrabec, J.; Stoll, J.; Hasse, H. A Set of Molecular Models for Symmetric Quadrupolar Fluids. *J. Phys. Chem. B* **2001**, *105*, 12126–12133.

(40) Brown, J. S. Interatomic potential parameters of solid neon and argon. *Proc. Phys. Soc.* **1966**, *89*, 987–992.

(41) Sengers, J. V.; Sengers, J. M. H. L. Thermodynamic Behavior of Fluids Near the Critical Point. *Annu. Rev. Phys. Chem.* **1986**, *37*, 189–222.

(42) Widom, B.; Rowlinson, J. S. New Model for the Study of Liquid–Vapor Phase Transitions. *J. Chem. Phys.* **1970**, *52*, 1670–1684.

(43) Mick, J. R.; Soroush Barhaghi, M.; Jackman, B.; Rushaidat, K.; Schwiebert, L.; Potoff, J. J. Optimized Mie potentials for phase equilibria: Application to noble gases and their mixtures with n-alkanes. *J. Chem. Phys.* **2015**, *143*, 114504.

(44) Neuefeind, J.; Feygenson, M.; Carruth, J.; Hoffmann, R.; Chipley, K. K. The Nanoscale Ordered MAterials Diffractometer NOMAD at the Spallation Neutron Source SNS. *Nucl. Instrum. Methods. Phys. Res. B.* **2012**, *287*, 68–75.

(45) Lawson, A. C.; Olsen, C. E.; Richardson, J. W.; Mueller, M. H.; Lander, G. H. Structure of Beta-uranium. *Acta. Crystallogr. B. Struct.* **1988**, *44*, 89–96.

(46) Wilson, A. S.; Rundle, R. E. The structures of uranium metal. *Acta Cryst.* **1949**, *2*, 126–127.

(47) Iosilevskiy, I.; Gryaznov, V. Uranium critical point problem. *J. Nucl. Mater.* **2005**, *344*, 30–35.

(48) Elrod, M. J.; Saykally, R. J. Many-body effects in intermolecular forces. *Chem. Rev.* **1994**, *94*, 1975–1997.

(49) Guillot, B.; Mountain, R.; Birnbaum, G. Theoretical study of the three-body absorption spectrum in pure rare gas fluids. *Mol. Phys.* **1988**, *64*, 747–757.

(50) Guillot, B. Triplet dipoles in the absorption spectra of dense rare gas fluids. II. Long range interactions. *J. Chem. Phys.* **1989**, *91*, 3456–3462.

(51) LeSar, R. Ground- and excited-state properties of solid argon under pressure. *Phys. Rev. B* **1983**, *28*, 6812–6820.

(52) Bafile, U.; Ulivi, L.; Zoppi, M.; Moraldi, M.; Frommhold, L. Third virial coefficients of collision-induced, depolarized light scattering of hydrogen. *Phys. Rev. A* **1991**, *44*, 4450–4458.

(53) Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.

(54) Sadus, R. J. Exact calculation of the effect of three-body Axilrod–Teller interactions on vapour–liquid phase coexistence. *Fluid Ph. Equilibria* **1998**, *144*, 351–359.

(55) Marcelli, G.; Sadus, R. J. Molecular simulation of the phase behavior of noble gases using accurate two-body and three-body intermolecular potentials. *J. Chem. Phys.* **1999**, *111*, 1533–1540.

(56) Soper, A. Inelasticity corrections for time-of-flight and fixed wavelength neutron diffraction experiments. *Mol. Phys.* **2009**, *107*, 1667–1684.

(57) Rosenberger, D.; Hanke, M.; van der Vegt, N. F. Comparison of iterative inverse coarse-graining methods. *Eur. Phys. J. Spec. Top.* **2016**, *225*, 1323–1345.

(58) Bernhardt, M. P.; Hanke, M.; van der Vegt, N. F. A. Iterative integral equation methods for structural coarse-graining. *J. Chem. Phys.* **2021**, *154*, 084118.

(59) Sivia, D. *Elementary Scattering Theory: For X-ray and Neutron Users*; Oxford University Press: New York, 2011.

(60) Allen, M. P.; Tildesly, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 2017.

(61) Louis, A. A. Beware of density dependent pair potentials. *J. Phys. Condens. Matter* **2002**, *14*, 9187.

(62) Barocchi, F.; Chieux, P.; Magli, R.; Reatto, L.; Tau, M. Static structure of dense krypton and interatomic interaction. *J. Phys. Condens. Matter* **1993**, *5*, 4299–4314.

(63) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.

(64) Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **1974**, *49*, 197–198.

(65) Zwicker, J.; Lovett, R. When does a pair correlation function fix the state of an equilibrium system? *J. Chem. Phys.* **1990**, *93*, 6752–6755.

(66) Frommer, F.; Hanke, M.; Jansen, S. A note on the uniqueness result for the inverse Henderson problem. *J. Math. Phys.* **2019**, *60*, 093303.

(67) Gray, C. G.; Gubbins, K. E. Theory of molecular fluids. Volume 1: Fundamentals. *Int. J. Quantum Chem.* **1990**, *38*.

(68) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids: with Applications to Soft Matter*; Academic Press: San Diego, 2013.

(69) Rasmussen, C. E.; Williams, C. K. I. *Gaussian processes for machine learning*; MIT Press: Cambridge, Mass, 2006.

(70) Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. *Bayesian Data Analysis*; Chapman and Hall/CRC: New York, 1995.

(71) Ambrogioni, L.; Maris, E. Integral Transforms from Finite Data: An Application of Gaussian Process Regression to Fourier Analysis. Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics. 2018; pp 217–225, ISSN: 2640-3498.

(72) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* **2020**, *173*, 109363.

(73) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(74) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. Python Sci.* **2016**, 98–105.

(75) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.

(76) Bellissent-Funel, M. C.; Buontempo, U.; Filabozzi, A.; Petrillo, C.; Ricci, F. P. Neutron diffraction of liquid neon and xenon along the coexistence line. *Phys. Rev. B* **1992**, *45*, 4605–4613.

(77) Yarnell, J. L.; Katz, M. J.; Wenzel, R. G.; Koenig, S. H. Structure Factor and Radial Distribution Function for Liquid Argon at 85K. *Phys. Rev. A* **1973**, *7*, 2130–2144.

(78) Barocchi, F.; Chieux, P.; Magli, R.; Reatto, L.; Tau, M. Neutron diffraction study of liquid krypton and the interatomic interaction. *Phys. Rev. Lett.* **1993**, *70*, 947–950.